

Proyecto Bi_Etapa_2

1. (20%) **Proceso de automatización del proceso de preparación de datos, construcción del modelo, persistencia del modelo y acceso por medio de API:**

Al momento de realizar el proceso de automatización y preparación de datos para la construcción del modelo, persistencia del modelo y acceso por medio de API se decidió se decidió que lo primero que se pensaba hacer era tener la limpieza de datos aparte, esto con el fin de tener más limpieza y orden a la hora de crear los pipelines y aplicarlos al API. Para la construcción del modelo del pipeline. Lo que se realiza es la importación de librerías. Luego con esto se realiza el pipeline. Como en la entrega pasada se menciono que se hicieron pruebas con los dos vectorizados, countVectorizer y TfidfVectorizer una vez teniendo eso se decidió usar TfidfVectorizer para los pipelines. Una vez se tenía esto se leen los datos, se dividen en prueba y entrenamiento, se hace el pipeline. Se sacan de igual manera las matrices de confusión las cuales nos sirvieron para ver como mejoraba el modelo con la persona de estadística. Para la persistencia del modelo se crea un joblib del pipeline para la velocidad y un pickle para la persistencia.

Por ultimo mediante el uso de Fast API, la cual se implemento y nos guiamos de esta página: <https://fastapi.tiangolo.com/es/>

1. Se usa @app.post("/predict/") el cual es un método que convierte los json en arreglos 2D y los guarda en una variable. Luego de esto hace un Excel con las predicciones y se hace run del main usando uvicorn en el puerto 8000.

2. (40%) **Desarrollo de la aplicación y justificación.**

En cuanto a la aplicación web lo que se definió en cuanto a los roles, es que nosotros los estudiantes (somos los desarrolladores de la aplicación). Las Naciones Unidas es la empresa la cual esta tomando el servicio y por ultimo el usuario es el trabajador de Naciones Unidas el cual va a usar la aplicación para facilitar y hacer de forma rápida y efectiva el trabajo que tiene en la compañía. En este caso es importante la existencia de la aplicación, ya que no todas las personas tienen los estudios con el fin de poder entender las matrices y o resultados obtenidos. Por este motivo es importante la aplicación ya que permite que cualquier persona, sin necesidad de una preparación antes pueda usarla de manera óptima. La forma de usarse es sencilla, se entra un excel con los textos o texto y este los clasifica. Gracias a la persona de estadística se pudo evidenciar una mejora en la matriz de confusión, lo cual representa una muy buena mejora para el modelo. Para esto se tuvo en cuenta los datos que

estábamos usando en el proyecto, como los tratábamos y limpiábamos. Como es que el usuario recibía los datos y se acordó que la mejor forma era por cada texto enviar únicamente el código 1,2,3,4,5 acorde a cada una de las necesidades. Se acordó el usuario de una mejor manera para acercarse a él. En cuanto al aporte de la persona de estadística se ayudo a la mejora de datos y bajar a la confusión del modelo es decir la capacidad de equivocarse del modelo.

3. (18%) Resultados.

los resultados se presentan en el video que esta en el paddlet.

4. (10%) Trabajo en equipo.

Líder de proyecto	Juan Diego Castellanos Bonilla
Ingeniero de datos	Carlos Perez
Ingeniero de software responsable del diseño de la aplicación y resultados	Ana Sofia Medina. (se aclara que como se trabajó en conjunto el video se graba por Tales Losada.)
Ingeniero de software responsable de desarrollar la aplicación final	Ana Sofia Medina. (se aclara que se trabajo en conjunto de Juan Castellanos)

Reuniones que se hicieron presentes:

Reunión de lanzamiento y planeación	<p>La reunión de la planeación y lanzamiento se realizó el lunes 8 de abril donde se definieron roles y forma de trabajo de grupo.</p> <p>En esta reunión se definió: Debido al buen trabajo en la etapa 1 del proyecto la líder seria: Ana Sofia Medina.</p> <p>En cuanto el tratado de datos: debido a las materias que vio y ve Carlos se definió el para mejorar, optimizar y mejorar el trabajo de los datos.</p> <p>En cuanto al diseño de aplicación web: debido a los trabajos realizados por Juan Diego Castellanos se decidió que el seria el ingeniero para trabajarla.</p> <p>por último la implementación de la Api se hizo en conjunto y el</p>
--	---

	desarrollo del pipeline lo realizo Juan Diego Castellanos como ingeniero, seleccionando el mejor modelo.
Reuniones de seguimiento	Las reuniones de seguimiento se hicieron el día 18 luego de sustentar la Etapa 1 con el fin de saber como estaban y si necesitaban ayuda, ya que una vez se tenía el API construida era necesario hacer el Font para mostrar datos.
Reunión con la persona de estadística	La reunión con la persona de estadística. Se hizo mediante WhatsApp ya que en este caso se comenzó la comunicación por correo. Una vez hizo esto se comenzó a hablar con ellas mediante WhatsApp debido a que las únicas horas disponible para ella era luego de las 4 de la tarde. Se decidió hablar y hacer preguntas por WhatsApp para poder saber que consideraría ella para cambiar el proyecto y mejorar las métricas.
Reunión de finalización	Se tiene programada la reunión a las 7 de la noche en la cual se va a revisar todo lo implementando y como funciona todo. Se grabara el video de la sustentación y se realizara la entrega del proyecto.

5. (12%) Sustentación y evaluación del aporte individual